NVM Express Technical Errata

Errata ID	012
Change Date	6/2/2011
Affected Spec Ver.	NVM Express 1.0a
Corrected Spec Ver.	

Submission info

Name	Company	Date
Peter Onufryk	IDT	5/11/2011
Kevin Marks	Dell	5/11/2011
Ken Okin	Virident	5/11/2011
Darren Lasko	Intel	5/11/2011

This erratum makes editorial changes to section 5.

This erratum clarifies that the default value for the interrupt Aggregation Time and Aggregation Threshold is 0h (disabled).

This erratum adds a maximum data transfer size to the Identify Controller data structure.

This erratum clarifies that Security Send and Security Receive utilize the Namespace Identifier field, as security is performed by the TCG on a namespace granularity.

Description of the specification technical flaw

Modify section 5.12.1.8 as shown below:

This Feature configures interrupt coalescing settings. This Feature applies only to the I/O Queues; interrupts are not coalesced for the Admin Queue. It is recommended that interrupts for commands that complete in error are not coalesced. The settings are specified in Command Dword 11.

The controller may delay an interrupt if it detects that interrupts are already being processed for this vector. Specifically, if the Completion Queue Head Doorbell register is being updated that is associated with a particular interrupt vector, then the controller has a positive indication that completion queue entries are already being processed. In this case, the aggregation time and/or the aggregation threshold may be reset/restarted upon the associated register write. This may result in interrupts being delayed indefinitely in certain workloads where the aggregation time or aggregation threshold are non-zero.

Figure 83: Interrupt Coalescing – Command Dword 11

Bit	Description
31:16	Reserved
15:08	**Aggregation Time (TIME):** Specifies the recommended maximum time in 100 microsecond increments that a controller may delay an interrupt due to interrupt coalescing. A value of 0h corresponds to no delay (i.e., disabling this capability). The controller may apply this time per interrupt vector or across all interrupt vectors. The reset value of this setting is 0h.
07:00	**Aggregation Threshold (THR):** Specifies the desired minimum number of completion queue entries to aggregate per interrupt vector before signaling an interrupt to the host; the default value is 0h. This is a 0's based value. The reset value of this setting is 0h.

Modify section 5.12.1.9 as shown below:

This Feature configures settings specific to a particular interrupt vector. The settings are specified in Command Dword 11.

By default, coalescing settings are enabled for each interrupt vector. Interrupt coalescing is not supported for the Admin Completion Queue.

Figure 84: Interrupt Vector Configuration – Command Dword 11

Bit	Description
31:17	Reserved
16	**Coalescing Disable (CD):** If set to '1', then any interrupt coalescing settings shall not be applied for this interrupt vector. If cleared to '0', then interrupt coalescing settings apply for this interrupt vector. By default, coalescing settings are enabled for each interrupt vector. Interrupt coalescing is not supported for the Admin Completion Queue.
15:00	**Interrupt Vector (IV):** This field indicates specifies the interrupt vector for which the configuration settings shall be are applied.

Modify section 5.12.1.10 as shown below:

This Feature controls write atomicity. The attributes are indicated in Command Dword 11.

Figure 85: Write Atomicity – Command Dword 11

Bit	Description
31:01	Reserved
00	**Disable Normal (DN):** If set to '1', then the host ~~indicates~~ specifies that the atomic write unit for normal operation is not required and that the controller shall only honor the atomic write unit for power fail operations. If cleared to '0', the atomic write unit for normal operation shall be honored by the controller.

Modify Figure 65 as shown below (takes ECN 008 changes into account):

Figure 65: Identify – Identify Controller Data Structure

Bytes	O/M	Description
		Controller Capabilities and Features
77	M	**Maximum Data Transfer Size (MDTS):** This field indicates the maximum data transfer size between the host and the controller. The host should not issue a command that exceeds this transfer size. If a command is processed that exceeds the transfer size, then the command is aborted with a status of Invalid Field in Command. The value is in units of the minimum memory page size (CAP.MPSMIN) and is reported as a power of two (2^n). A value of 0h indicates no restrictions on transfer size. The restriction includes metadata if it is interleaved with the logical block data.
255:~~77~~ 78		Reserved

Modify Figure 25 as shown below:

Figure 25: Opcodes for Admin Commands – NVM Command Set Specific

Opcode (07) Generic Command	Opcode (06:02) Function	Opcode (01:00) Data Transfer	Opcode	O/M	Namespace Identifier Used[3]	Command
1b	000 00b	00b	80h	O	Yes	Format NVM
1b	000 00b	01b	81h	O	~~No~~ Yes	Security Send
1b	000 00b	10b	82h	O	~~No~~ Yes	Security Receive

NOTES:
1. O/M definition: O = Optional, M = Mandatory.
2. Opcodes not listed are reserved.
3. A subset of commands uses the Namespace Identifier field (CDW1.NSID). When not used, the field shall be cleared to 0h.

Disposition log

5/11/2011	Erratum captured.
5/18/2011	Changed "power on value" to "reset value".
5/20/2011	Updated maximum data transfer size to indicate error value if exceeded.
5/26/2011	Minor editorial change to Maximum Data Transfer Size wording.
6/2/2011	Editorial changes to Maximum Data Transfer Size wording.
7/8/2011	Erratum ratified.

Technical input submitted to the NVMHCI Workgroup is subject to the terms of the NVMHCI Contributor's agreement.